DOCUMENT RESUME

ED 227 149 TM 830 160

AUTHOR Prater, Doris L.
TITLE What Counts as Effective University Teaching: The State of the Art.
PUB DATE Jan 83
NOTE 12p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Houston, TX, January 27-29, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Information Analyses (070)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *College Faculty; *Evaluation Methods; Higher Education; Merit Rating; Self Evaluation (Individuals); Student Evaluation of Teacher Performance; Summative Evaluation; *Teacher Effectiveness; *Teacher Evaluation; Teacher Promotion; Teacher Responsibility; Tenure

ABSTRACT

The appropriate questions in teacher evaluation in higher education are: who should evaluate, for what purposes, and using what means. Frequently, evaluation systems provide information for administrative decisions on promotion/tenure and merit. Increased reliance on student ratings of teachers has the disadvantage of distorting factors that may influence ratings, such as class size, lecture or discussion structure, and years teaching. When used as a measure of student satisfaction and supplemented with other sources, student ratings may be justified to judge some, but not all, aspects of teaching. Evaluating teacher effectiveness in terms of student achievement depends on whether the responsibility of learning rests with the pupil or the teacher. In addition to lack of agreement between faculty self-ratings and ratings by others, teachers tend to give themselves better ratings than their students do. Colleague evaluation by direct observation is best used formatively to improve teaching. Indirect review of syllabi, tests, and materials is generally useful with other evidence. The problems of defining effective teaching make evaluation particularly difficult. Evaluation will remain judgmental, but the best systems seem to use a variety of sources of evidence and explicate as clearly as possible how judgments will be made. (CM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED227149

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

WHAT COUNTS AS EFFECTIVE UNIVERSITY TEACHING:

THE STATE OF THE ART

Doris L. Prater

University of Houston at Clear Lake City

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D. Prater

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

TM 830 160

*Paper presented at the annual meeting of Southwest Educational Research Association, Houston, Texas, January 27-29, 1983.

WHAT COUNTS AS EFFECTIVE UNIVERSITY TEACHING: THE STATE OF THE ART

Background

Unlike scholarship, which has tangible evidence such as books and articles, results of effective teaching are not always immediately observable, are difficult to measure, and are not always attributable to a given source. Nevertheless, we are measured on how we teach in higher education and probably should be. As Millman says in the introduction to Handbook on Teacher Evaluation, "To evaluate or not to evaluate, that is NOT the question." (1981, p.12) Rather, he goes on to comment, the appropriate questions to ask are who should evaluate, for what purposes, and using what means. As to the purpose of evaluation, this discussion will focus for the most part on summative evaluation of teaching. That is, we will attend to evaluation systems that provide information for administrative decisions on such things as promotion/tenure and merit. And, we will consider the range of "whos" and "by what means".

Presently, the literature indicates that student ratings, student outcomes, colleague input, and self evaluation are the most often cited sources of evidence for these kinds of judgements. Each source has its own problems and, in practice, a combination of sources is often used, providing at least some system of checks and balances.

Surveys of evaluation practices (Astin and Lee, 1967; Riggs, 1975; Seldin, 1978; Muse, 1979; Centra, 1979) provide insights into current practice. Centra, reporting on a survey done in 1976 of 453 department heads of 134 colleges and universities reported increasing reliance on

student ratings and content of course syllabi and examinations. Department chairs ranked fifteen sources of information listed as most influential. Highest rankings went to chairman evaluations, systematic student ratings, and colleagues opinions. Least used sources of evidence about effective teaching were videotapes, long term follow-ups of students, alumni ratings, colleague ratings based on classroom visits, and student examination performance. When contrasted to surveys of evaluation practices in the 1960's reported by Astin and Lee, there is an increased reliance on student ratings. Such an increase was noted by Seldon (1978) who surveyed deans in over 400 colleges in 1973 and again in 1978. He found that in 1978, 53% "Always Used" student ratings, up from 29% in 1973. Riggs (1975) sampled 200 institutions and determined that 86% conducted student evaluations of teachers and in two-thirds of the cases, results were made available to administrators. Muse (1979) surveyed 53 state and 47 private teacher education programs in 14 states and found 84% used some kind of standardized instrument to rate faculty and 67% made the use of the instrument compulsory. With these findings in mind, let us look more specifically at the source of evidence most widely used, the student rating instrument.

Student Ratings

Student ratings are widely used. Generally, content of such forms includes broad areas dealing with organization and structure, teacher-student interaction or rapport, and skill in communicating and lecturing.(Centra, 1979) Despite the wide use of such rating forms, many disadvantages of their use are pointed out in the literature.

One often cited disadvantage of student ratings is the presence

of distorting factors that may influence ratings. Centra (1979) summarized the following course and teacher characteristics that have been shown to affect ratings:

1. Small classes of less than 12 students get especially high ratings and are also less reliable in their ratings. (Centra and Creech, 197 .)
2. Discussion classes get higher ratings than lecture classes. (E.T.S., 1975)
3. Students tend to give higher ratings to courses that are major requirements or electives as compared to courses that fulfill a college requirement.
4. Teachers are rated poorest in the first year of teaching. Teachers in years 3 - 12 did best. There was a slight decline in ratings for teachers with over 20 years experience. (Centra,1976)
5. No difference was found between academic rank of professors involved in above study. Teachers with teaching loads of 13+ hours received slightly higher ratings. This may be because they are generally located in two or four year colleges where teaching is the major activity. (Centra and Creech, 1976)
6. Research productivity, although seen as keeping teachers up to date in field in general did not relate to ratings received. (Aleamoni and Yimer, 1973)

Rotem and Glasman (1979) in reviewing effectiveness of student feedback in Review of Educational Research, Summer 1979, suggests that extraneous factors such as sex and rank of instructors, class size, class level, and expected grade have been found to be influential in some studies but inconsequential in others.

Another objection to the use of student ratings for evaluating college teaching, in addition to the distorting factors mentioned, is the credibility of students as valid sources of information. It has

been suggested that students are not competent judges of instruction because they lack experience, knowledge, and their judgement is affected by other factors. Marsh (1977) suggests that students may have implicit theories about teaching and that the ratings tell us more about the students and what they consider ideal teaching than about the teacher being rated. Levinthal, Abrami, and Perry (1976) suggest that students do not all enter courses with the same expectations and therefore ratings of teacher performance are based on different reference points and probably ought not be aggregated.

Looking at the research in general, a criterion-related approach to establishing the validity of student ratings has most often been used. Relationships are demonstrated between student ratings and such things as colleagues ratings, administrators ratings, and most frequently with student achievement.

Cohen (1981) synthesized the research on the relationship betw student ratings of instruction and student achievement using meta analysis techniques in Review of Educational Research, Fall 1981. He analyzed 41 independent studies. The average correlation between overall instructor ratings and student achievement was .43; between overall course rating and achievement was .47. The author concludes that results appear to support validity of student ratings as measures of teaching effectiveness.

Centra (1977) also cites strong positive correlations between student ratings and achievement. Rodin and Rodin (1972) report one of the few findings of negative relationship. However, the methodology in this study has been questioned, most notably because of the small sample and nature of the instruction provided.

In general, reviewers acknowledge small to moderate positive relationships between ratings and learning. However, a number of problems exist with measuring student achievement such as ways to assess and how to control for initial ability. These problems cast some doubt on exactly what we are accomplishing by validating one measure (student ratings) with its inherent problems with another measure (student achievement) with its inherent problems. Both measures may be invalid. Marsh (1980) suggests that since there is a lack of unanimity on a definition of good teaching that we might more appropriately look at establishing construct validity. Specifically, he suggests that we continue to look for effectiveness criteria and for positive correlations with those factors assumed to be related, negative correlations with those factors assumed to be unrelated.

In summary, although widely used, student ratings present some problems. However, when used merely as a measure of student satisfaction and when supplemented with other sources of evidence, they may be justified. Students may be qualified to judge some aspects of teaching, although certainly not all aspects.

Student Achievement

McKeachie (1971) and others suggest that a criteria of teaching effectiveness ought to include student achievement. According to Travers (1981), the difficulties of assessing teacher effectiveness in terms of test scores of pupils , although commendable on the face of it, present almost insurmountable problems. There is, first of all,

the question as to whether the responsibility of learning should rest with the pupil or with the teacher. Over the past half century, there has been a slow accumulation of studies that have attempted to discover relationships between pupils measured gains in achievement under a particular teacher and personal characteristics of the teacher. (e.g.,Rosenshine and Furst's work) Such studies generally show the kind of relationship that one would expect on the basis of common sense. But, he goes on, research in the last decade has tended to move away from studying the teacher and toward studying the pupil. (e.g., time spent on learning, Hoge and Luce, 1979).

Millman (1981) points out that many factors impact on student achievement including teacher performance, the particular measure of achievement used, and characteristics of the student including prior learnings and learning outside the teacher setting. Scriven (1981) points out that the best teaching is not that which produces the most learning since that which is learned may be useless . He indicates that there is a connection between the two, but not a direct one. Scriven, incidentally suggests using student ratings as a key component. Within the questionnaire, he sees the crucial question as a simple request for an overall judgement of merit of the instructor. He would also use peer or expert review of teaching materials, exit interviews, and professional development dossiers.

Self-Evaluation

Blackburn and Clark (1975) reported little agreement between faculty self-rating of overall teaching and ratings by students, colleagues, or administrators. In addition to lack of agreement,

Centra (1973) reports a tendency for teachers to give themselves better ratings than their students do. In comparing ratings on 343 teachers (self and student) from five colleges, he found the most discrepancies between student-teacher interactions, course objectives, and instructors openess to others viewpoints.

Colleague Evaluation

Colleague evaluation can be divided into direct observation of teaching by colleagues or the more indirect review of teaching materials, syllabi, tests, etc. When rating teaching, colleagues tend to be generous. Centra (1975) reported 94% of teachers in a large study were rated "excellent" or "good" by colleagues. This rating was attained after each teacher was observed twice by three observers or a total of six times. Teachers in the study also tended to be inconsistent on many dimensions of the rating scale as well.

Scriven (1981) gives several reasons not to use classroom visits to evaluate teaching. First, the visit itself alters the teaching and a representative sample is not observed. Second, the number of visits is generally too small to be an accurate sample. Third, the visitors are not generally devoid of personal feelings toward the teacher. Fourth, nothing observed in the classroom can be used as a basis for an inference to any conclusion about the merit of teaching because style indicators have not been shown to correlate with short or long term learning of students. In general, classroom observations by colleagues are probably best used formatively to improve teaching and by mutual consent of observers and observee.

Colleague's review of teaching documents such as syllabi and tests is fairly widely used. (Centra, 1979, Scriven, 1981) When used with other sources of evidence, such information is generally viewed as useful.

Summary

Research to date has not produced a clear direction for evaluating teaching in higher education. We are still struggling with problems of definition of effective teaching which make the task of evaluation particularly difficult. Decisions about the quality of one's teaching will remain judgemental for the most part. However, the best systems seem to use a variety of sources of evidence and attempt to explicate as clearly as possible how judgements will be made.

References

Aleamoni, L. M., and Yimer, M. An investigation of the relationship between colleague ratings, student ratings, research productivity, and academic rank in rating instructional effectiveness. Journal of Educational Psychology, 1973, 64, 274-277.

Astin, A. W., and Lee, C. T. T. Current practices in the evaluation and training of college teachers. In C. B. T. Lee (Ed.), Improving College Teaching, Washington, D. C.: American Council on Education- 1967.

Blackburn, R. T. and Clark, M. J. An assessment of faculty performance: Some correlates between administrator, colleague, student and self-ratings. Sociology of Education, 1975, 48, 242-256.

Centra, J. A. Self-ratings of college teachers: A comparison with student ratings. Journal of Educational Measurement, 1973, 10, 287-295.

Centra, J. A. Determining Faculty Effectiveness. San Francisco: Jossey-Bass Publishers, 1979.

Centra, J. A. Student ratings of instruction and their relationship to student learning. American Educational Research Journal, 1977, 14, 17-24.

Centra, John. (Ed.) Renewing and Evaluating Teaching. San Francisco, Ca.: Jossey-Bass Publishers, 1977.

Centra, J. A. and Creech, F. R. The relationship between student teachers, and course characteristics and student ratings of teacher effectivenss. Project Report 76-1. Princeton, N. J.: Educational Testing Service, 1976.

Druger, M. On the evaluation of college teaching. Journal of College Science Teachers, 1982, 11, 254-55.

Hoge, R. D. and Luce, S. Predicting achievement from classroom behavior. Review of Educational Research, 49, 479-496.

Levinthal, L., Abrami, P. C., and Perry, R. P. Do teacher rating forms reveal as much about students as about teachers? Journal of Educational Psychology, 1976, 68, 441-445.

Ley, R. Tenure and student evaluations of teaching. College Student Journal, Spring 1981, 15, 147-150.

Marsh, H. W. The validity of students' evaluation: Classroom evaluations of instructors independently nominated as best and worst teachers by graduating seniors. American Educational Research Journal, 1977, 14, 441-447.

Marsh, H. W. and Overall, J. U. Validity of students' evaluations of teaching effectiveness: Cognitive and affective criteria. Journal of Educational Psychology, 1980, 72, 468-475.

Marsh, H. W., Validity of students' evaluations of college teaching: A multi-trait multi-method analysis. Journal of Educational Psychology, 1982, 72, 264-279.

McKeachie, W. J. Lin, Y. and Mann, W. Student ratings of teacher effectiveness: Validity studies. American Educational Research Journal, 1971, 8, 435-445.

Millman, J. (Ed.) Handbook of Teacher Evaluation. Beverly Hills, Ca.: Sage Publications, 1981.

Muse, Corey. Faculty evaluation. Improving College and University Teaching, Winter 1979, 44.

Renner, Richard. Comparing professors: How student ratings contribute to the decline in quality of higher education. Phi Delta Kappan, October, 1981, 128-130.

Riggs, Robert. The prevalence and purposes of student and subordinate evaluations among AACTE member institutions. Journal of Teacher Education, Fall, 1975, 119.

Rodin, M. and Rodin, B. Student evaluations of teachers. Science, 1972, 177, 1, 1164-1166.

Rotem, R. and Glasman, N. Effectiveness of students' feedback. Review of Educational Research, 1979, 49, 497-511.

Scriven, M. Summative teacher evaluation. In J. Millman (Ed.) Handbook of Teacher Evaluation. Beverly Hills, Ca.: Sage Publications, 1981

Travers, Robert M. W. Criteria of good teaching. In J. Millman (Ed.) Handbook of Teacher Evaluation. Beverly Hills, Ca.: Sage Publications, 1981.